Assessing the feasibility of developing a four-dimensional (4-D) interpolator for use in impaired waters listing assessment

December 2008
STAC Publication 08-008

Recommendations from the STAC Expert Panel
Frank Curriero (Johns Hopkins University)
Eileen Hofmann (Old Dominion University)
Ragu Murtugudde (University of Maryland)
Jian Shen (Virginia Institute of Marine Science)
J. Andrew Royle (U.S. Geological Survey)

Preamble

The Chesapeake Bay Program (CBP) has been collecting environmental data (e.g. temperature, dissolved oxygen) at monitoring sites throughout the Bay since 1985. These data provide the potential for developing data products that can be used to inform and guide water quality decisions and policies for Chesapeake Bay. Thus, several groups within the Chesapeake Bay community have expressed an interest in a four-dimensional (4-D) interpolator, and/or have begun efforts to develop one.

To obtain guidance and inputs on how to develop a 4-D interpolator and to identify approaches for its use with the Bay monitoring data, the Chesapeake Bay Scientific and Technical Advisory Committee (STAC) convened an expert panel. This panel met on 10 December 2007 at the Chesapeake Bay Office in Annapolis, MD. The morning session consisted of a workshop with presentations to review possible approaches for integrating and synthesizing the Bay data sets to provide products that are useful for water quality decisions. The afternoon discussions focused on developing an approach for moving forward with analyses of the Bay environmental data sets. The overall purpose of the workshop and panel meeting was to identify a core set of functionality requirements and to facilitate the coordination of development efforts now ongoing by several groups, so that future products will meet the needs of CBP partners.

This report is a summary of the discussions that ensued at the workshop and the expert panel discussions. It provides the recommendations from these discussions.

Background

What is meant by 4-D interpolation?

A 4-D interpolation is taken here to represent the process by which data are interpolated (or predicted) in four dimensions. For the Chesapeake Bay applications that are subsequently described these dimensions include the spatial indices, longitude, latitude, and depth, and a temporal index. Data available to support 4-D interpolation will be indexed by these dimensions and are taken generically to represent observed measurements or output from other products, such as simulations from water quality models and data simulation/assimilation schemes. The process of 4-D interpolation will likely develop from a combined set of scientific techniques. The integrated product (software, protocols, etc.) that performs the 4-D interpolation is termed the 4-D interpolator.

Why is the Chesapeake Bay Program interested in 4-D Interpolation?

The numerical water quality criteria established by Chesapeake Bay Program partners for the 303d listing process require assessment at space and time scales that are at higher resolution than is feasible with currently available data. For example, the water quality criteria require Bay-wide assessment of dissolved oxygen as a 30-day mean, a 7-day mean, a 1-day mean, and an instantaneous measure. However, long-term fixed station monitoring data (e.g. nutrients, dissolved oxygen, salinity, and temperature) are collected only once or twice per month, in open and deep waters, at fixed points in space, and therefore lack the appropriate temporal and spatial resolution to meet these criteria. Other projects, such as the Continuous Monitoring Program and the DATAFLOW monitoring cruises, collect these data more frequently at higher spatial resolution, but only for limited time periods and over limited spatial extents. For example, data from the continuous monitoring programs in Maryland and Virginia contain very fine temporal resolution, but they are restricted to a limited number of shallow water habitats.

The development of an approach(es) that allows 4-D interpolation of the monitoring data sets to the entire Chesapeake Bay and its tidal tributaries would allow combining and integrating measurements from numerous disparate datasets to generate a more complete interpolation of available data in space *and* time. This would improve the ability to evaluate water quality for the 303d listing process.

Current estimates of nutrient and sediment fluxes into the tidal portions of the Bay are constrained by the ability to efficiently integrate data that have been collected at different scales in space and time. A 4-D Interpolator would enhance the ability to estimate such fluxes by providing fields that can then be combined with circulation fields determined from either direct measurements or numerical circulation models. It could also help to spatiotemporally optimize the application of selecting the best management practices for reducing point and nonpoint source pollution.

Recent studies suggest that Chesapeake Bay has become more sensitive to nitrogen (N) loading in the past 20 years, with higher volumes of hypoxic volume being reported in response to similar N loads (Hagy et al. 2004). However, the ability to accurately estimate the temporal and spatial extent of hypoxia in the Bay is limited by current sampling density, monitoring instrumentation, and analytical tools. A 4-D interpolation protocol would enable better targeting of the placement of new monitoring tools, such as automated sensors, in locations that optimize the utility of the data being collected.

Functional Requirements of a 4-D Interpolator

An output from the STAC-sponsored workshop and discussions was identification of requirements for using a 4-D interpolator for listing assessment. This was considered to be critical to defining the approach(s) to be used. These requirements are as follows:

1. The interpolated fields should allow evaluation of a 1-day mean, a 7-day mean, and a 30-day mean. Because of the importance of tidal and advective circulation motions for dissolved oxygen variability, temporal resolution of the data needs to be at least two measurements per tidal cycle.

2. Interpolated values should be accompanied by statistical estimates of uncertainty.

3. Functionality must exist to automate the interpolation process, eliminating the need for subjective "expert" decisions that are manually implemented at various stages of interpolation. Expert knowledge will go into developing, maintaining, and updating the interpolation scheme.

4. Interpolation must be un-biased with respect to the data and capable of "filling in the gaps" between monitoring data.

5. Interpolated parameters include at least dissolved oxygen, and preferably chlorophyll a concentrations and water clarity.

6. The interpolator must recognize land boundaries and not interpolate across land.

7. The interpolation method adopted by the Bay Program should be based on approaches with a sound scientific basis that are supported by studies in the peer-reviewed scientific literature.

8. The properties of the chosen interpolation method that identify it as the optimal method for the intended application should be characterized.

Feasibility of a 4-D Interpolator

After defining the requirements for a 4-D interpolator, the panel next discussed the feasibility of identifying an interpolator protocol that could satisfy the above requirements. For this application, feasibility was defined as, "*identification of a known method or set of methods that can be applied to the existing Bay environmental data sets using the current state of knowledge about processes in the Bay*". The evaluation of the feasibility of a 4-D interpolator for the Bay Program was focused around 1) availability of sufficient data, and 2) existence of suitable interpolation methods. These discussions are summarized as follows.

Evaluation of Existing Datasets:

A consensus opinion from the expert panel was that the sampling frequency and spatial resolution of the existing Chesapeake Bay datasets are insufficient for successful extrapolation to four dimensions. However, there is an on-going effort among Chesapeake Bay partners to acquire funding to deploy continuous monitoring buoys, which are equipped with vertical profilers in deep water areas of the Chesapeake Bay and tidal tributaries. If these efforts succeed, then the shortcomings of existing datasets will be greatly alleviated.

In an effort to better understand issues regarding data sufficiency, the expert panel recommended evaluation of the existing datasets using a range of analytical techniques to develop our understanding of key properties of the existing datasets. This information will be critical in supporting development and evaluation for any 4-D interpolator procedure that is chosen. Such analyses can be started now and do not need to wait for the 4-D interpolator scheme to be developed. Suggestions for these analyses are:

1. implement processing of the long-term Chesapeake Bay dataset to develop metadata that provides information on sample frequency, e.g. phase of tide when data were acquired,
2. estimate autocorrelation scales for selected fields, such as oxygen. This can include assessments of anisotropy in the spatial and/or space-time dimensions (Cressie 1991, Schabenberger and Gotway 2004, Diggle and Ribeiro 2006), knowledge of which would be crucial for supporting a kriging based 4-D interpolation approach (see Appendix) and other methods,
3. use kriging methods to assess output from the Chesapeake Bay eutrophication model, in order to test optimal sampling frequency and optimal sampling locations for data that would be input to a 4-D interpolator. Sampling the simulated data in various ways and using information on autocorrelation scales (item 2) will provide insights into variations of the data covariances and error propagation,
4. estimate potential bias in measurements, such as might occur via sampling strategy,
5. undertake a simple bilinear analysis,
6. undertake an analysis of long-term trends in the Bay data sets using an approach such as wavelet analysis or other time series techniques,
7. undertake an empirical orthogonal function analysis to determine dominant patterns in the data sets,
8. investigate internal or non-deterministic variability,
9. perform a signal/noise calculation for the observations (using an approach similar to Ballaberra et al. 2003),
10. undertake some small-scale data assimilation experiments to determine the frequency and types of data that are needed to develop the larger scale fields that are needed. Simple data assimilation procedures such as nudging or Kalman filtering and smoothing (Stauffer and Seaman, 1990; Evensen, G. 1992) provide approaches that could be implemented with existing data sets. Complex data assimilation approaches, such as variational methods, are not feasible given the

current state of knowledge about the Bay processes and current state of analysis of the Bay data sets.

The analyses listed above are recommended for the task of assessing data sufficiency. In addition, the expert panel is recommending several approaches that might provide a promising basis for developing a 4-D interpolator. These are briefly described in the Appendix.

Adequate Interpolation Methods:

A recommendation regarding the adequacy or optimality of any particular interpolation method first requires conducting exploratory studies of different interpolation approaches. The expert panel recommended that any interpolation method should satisfy the following criteria:

- Scientifically Defensible: The chosen interpolation method should be based on accepted, scientifically-based approaches and be appropriate for the intended application.
- Method Evaluation: the interpolation method should include approaches for evaluation of uncertainty and errors of the interpolated fields.
- Method Implementation: the implementation of the interpolation method must be objective and scientifically defensible.
- Useable as a Tool: the interpolation method should be flexible, robust and capable of implementation in an automated mode.

Summary

The consensus of the expert panel was that there is currently insufficient information to evaluate the feasibility of a 4-D interpolator for use in water quality assessment or related activities.

The panel recommended that a study be initiated to evaluate the different approaches available for developing a 4-D interpolator, that includes evaluation of optimal sample locations and sampling frequency.

The panel further recommended that data analysis studies be initiated to develop the statistical basis for a 4-D interpolator.

APPENDIX: Potential Interpolation Methods:

In the process of preparing this report, the expert panel discussed several methods that hold promise for developing a 4-D interpolator that can meet the functional requirements outlined above. This section provides summaries of the most promising methods. The intent of this appendix is to provide a brief explanation of each method and indicate what it can contribute to 4-D interpolation of Chesapeake Bay data sets. References are included as a starting point for investigating these methods. Inclusion in this list does not imply endorsement of a particular method. Rather, the expert panel recommended that all listed methods be investigated as part of the eventual development of a 4-D interpolator for Chesapeake Bay

Extending Kriging for Bay Applications to the 3-D and 4-D Scenarios.

Kriging is a spatial interpolation technique that arose out of the field of geostatistics, a subfield of statistics that deals with the analysis of spatial data. Kriging and the field of geostatistics have been employed in a wide variety of environmental applications and is generally accepted as a method for performing statistically optimal spatial interpolations (Cressie 1991, Schabenberger and Gotway 2004, Diggle and Ribeiro 2006). Applications of kriging in water related research are found in (Kitanidis 1997, Wang and Liu 2005, Ouyang et al. 2006) with some specific to Chesapeake Bay applications (Chehata et al. 2004, Jensen et al. 2006,Chehata et al. 2006). Details on kriging methodology, geostatistics, and their related statistical development can be found in (Cressie 1991, Diggle et al. 1998, Schabenberger and Gotway 2004, Diggle and Ribeiro 2006).

There are several issues regarding the optimal use of kriging for Bay-specific applications in the less complicated 2-D space that should first be well understood and researched before moving to applications in 3-D and 4-D space. Currently available data are sufficient to work through these details regarding the use of kriging for Chesapeake Bay applications. Secor et al. (2006) provide a detailed overview on the application of kriging to spatial interpolation of water quality parameters in Chesapeake Bay.

Hindcasting

Hindcast experiments can be conducted in order to validate kriged products. This type of approach is useful when combined with a quantitative measure of skill score, such as Taylor diagrams (Taylor 2001) or target diagrams (Joliff et al. in press).

Data Assimilation: The "Optimal Interpolation" Method in combination with Kriging

The optimal interpolation method, currently implemented in the Regional Ocean Modeling System (ROMS), could be a start towards testing data in a relatively small domain. The term "optimal interpolation" is used here to refer to a specific data assimilation method identified by that name, and has no implications across other methods considered. In other words, this does not imply that the optimal interpolation method has been determined to be "optimal" for the development of a 4-D interpolator.

There is a version of ROMS currently implemented for Chesapeake Bay. The method uses fixed covariances, which is similar to what is used with kriging (Kantha and Clayson, 1994; Buehner and Malanotte-Rizzoli)[1]. Combined with the knowledge obtained from a kriging method, this will provide a more robust approach for specifying the covariance. Kriging can also be used to blend model-derived fields with observational data, where available. If using relatively strong temporal weighting (Cheng et al., 2004), the correction information can propagate to the next time interval when new observational data become available. In this way, daily and weekly interpolation might be feasible. It may be useful to evaluate the possibility of using output from a long-term simulation of the Chesapeake Bay eutrophication model as input to an interpolator. This will provide a test of the skill and accuracy of the interpolation routine.

Neural Networks

Neural networks use the same input as kriging and could be used to develop an interpolation method based on neural nets. Neural networks have been applied to the prediction of water quality parameters (Maier and Dandy 1996; Schulze et al. 2005). This may also help to reduce the computation time of dynamical models. A neural net for the equation of state significantly reduced computational time in the NCEP ocean model (Krasnopolsky and Chevallier 2003).

Implementing Land Boundaries

Grid generators could be used to provide land boundaries. The Bay data are distributed along tributaries and in the main stem making calculation of gradients problematic. The dynamical models are already using unstructured grids or orthogonal curvilinear grids that cover the Bay and give indices to track gradients accurately once the data are distributed on the grid. Note that the data do not need to be gridded. Rather, the grid indices can be used to locate the data values in an array and the gradients can then use the array and the indices.

---

[1] In kriging, covariances are first estimated using available data, and then assumed to be fixed for interpolation purposes

References

Ballabrera-Poy, J., R.G. Murtugudde, J.R. Christian, and A.J. Busalacchi. 2003. Signal-to-noise ratios of observed monthly tropical ocean color. *Geophysical Research Letters* 30(12): 1645.

Buehner, M., and Malanotte-Rizzoli, P. 2003. Reduced-rand Kalman filters applied to an idealized model of the wind-driven ocean circulation. Journal of geophysical Research, 108(C6), 3192-3215.

Chehata, M., Jasinski, D., Monteith, M., and Samuels, W.B. 2007. Mapping Three-Dimensional Water Quality Data in the Chesapeake Bay Using Geostatistics. , *Journal of the American Water Resources Association*. 43(3):813-826.

Chen, C., Beardsley, R.C., Cowles, G. 2004. An Unstructured Grid, Finite-Volume Coastal Ocean Model FVCOM User Manual. University of Massachusetts-Dartmouth, pp184.

Courtier, P., Thepaut, J.N., Hollingsworth, A. 1994. A strategy for operational implementation of 4DVAR, using an incremental approach, Q.J.R. Meteorol., Soc. 120,1367-1387.

Cressie, N, 1991. Statistics for Spatial Data, Wiley, New York, 928 p.

Diggle, PJ, Tawn, JA, and Moyeed, RA, 1998. Model Based Geostatistics (with Discussion), Applied Statistics, v. 47, p. 299-350.

Diggle, PJ and Ribeiro PJ, 2006. Model-based Geostatistics, Springer, New York, 230p.

Evensen, G. 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. J Geophys Res 97(C11): 17 905–17 924

Hagy, J.D., W.R. Boynton, C.W. Keefe, and K.V. Wood. 2004. Hypoxia in Chesapeake Bay, 1950-2001: long-term change in relation to nutrient loading and river flow. *Estuaries* 27(4): 634-658.

Joliff, J., J.C. Kindle, I. Shulman, B. Penta, M.A.M. Friedrichs, R. Helber and R.A. Arnone. 2008. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. *Journal of Marine Systems* (in press).

Kantha, L. and C. A. Clayson, 1994. An improved mixed layer model for geophysical applications. J. Geophys. Res., 99 (C12), 25,235-25,266.

Kitanidis, PK, 1997. Introduction to Geostatistics: Applications in Hydrogeology, Cambridge University Press, New York, 271p.

Maier, H.R. and Dandy, G.C. 1996. The use of artificial neural networks for the prediction of water quality parameters. Water Resources Research. 32(4):1013-1022

Ouyang Y, Zhang JE, and Ou LT, 2006. Temporal and spatial distributions of sediment total organic carbon in an estuary river, *Journal of Environmental Quality,* 35, 93-100.

Secor, D., M Christman, F. Curriero, D. Jasinski, E. Perry, S. Preston, K. Reckhow, and M. Trice. 2006. The cumulative frequency diagram method for determining water quality attainment: report of the Chesapeake Bay Program STAC Panel to review Chesapeake Bay Program analytical tools. *STAC Publication 06 - 9 October 2006* (Available: http://www.chesapeake.org/stac/Pubs/CFD_STAC_Final.pdf).

Schabenberger, O and Gotway, CA, 2005. Statistical Methods for Spatial Data Analysis, Chapmann and Hall/CRC Press, Florida, 512 p.

Schulze, F.H., H. Wolf, H.W. Jansen, and P. van der Veer. 2005. Applications of artifical neural networks in integrated water management: fiction or future? Water Science and Technology 52(9): 21-31.

Stauffer, D. R., and Seaman, N. L., 1990. Use of four-dimensional data assimilation in a limited-area mesoscale model. Part 1: Experiments with synoptic-scale data. Mon. Wea. Rev. 118:1250-1277.

Taylor, K.E. 2001. Summarizing multiple aspects of model performance in a single diagram. *Geophysical Research Letters* 106:7183-7192.

Wang XJ and Liu RM, 2005. Spatial analysis and eutrophication assessment for chlorophyll a in Taihu Lake, *Environmental Monitoring and Assessment,*101, 167-174.